

Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes

Reshanne R. Reeder · Wieske van Zoest ·
Marius V. Peelen

© The Psychonomic Society, Inc. 2015

Abstract Theories of visual search postulate that the selection of targets amongst distractors involves matching visual input to a top-down attentional template. Previous work has provided evidence that feature-based attentional templates affect visual processing globally across the visual field. In the present study, we asked whether more naturalistic, category-level attentional templates also modulate visual processing in a spatially global and obligatory way. Subjects were cued to detect people or cars in a diverse set of photographs of real-world scenes. On a subset of trials, silhouettes of people and cars appeared in search-irrelevant locations that subjects were instructed to ignore, and subjects were required to respond to the location of a subsequent dot probe. In three experiments, results showed a consistency effect on dot-probe trials: dot probes were detected faster when they appeared in the location of the cued category compared with the non-cued category, indicating attentional capture by template-matching stimuli. Experiments 1 and 2 showed that this capture was involuntary: consistency effects persisted under conditions in which attending to silhouettes of the cued category was detrimental to performance. Experiment 3 tested whether these effects could be attributed to non-attentional effects related to the processing of the category cues. Results showed a consistency effect when subjects searched for category exemplars but not when they searched for objects semantically related to the cued category. Together, these results indicate that attentional templates for familiar object categories affect visual processing across the visual field, leading to involuntary attentional capture by template-matching stimuli.

Keywords Visual search · Attentional capture · Attention

Introduction

An everyday street scene contains many objects: cars on the road, other people on the sidewalk, streetlights, shop windows, etc. The visual system is limited in its capacity to process this information, so we must select those objects that are relevant for current goals for further processing. For example, if we want to cross the street, we have to attend selectively to cars (task-relevant) and temporarily ignore other objects. What are the mechanisms that allow us to efficiently perform tasks like these?

Theories of visual search postulate that the selection of targets amongst distractors involves matching visual input to a top-down attentional set or “search template.” The search template guides visual attention to items in the environment that contain target features, thereby facilitating target detection (Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989). In the case of real-world category search, such as looking for cars before crossing the street, an effective template must be composed of features that generalize to most target category exemplars but are specific enough to exclude non-target objects. Previous studies have found evidence that the most diagnostic features of a category are shapes of diagnostic object parts (e.g., wheels of a car; Evans & Treisman, 2005; Reeder & Peelen, 2013; Treisman, 2006; Ullman, Vidal-Naquet, & Sali, 2002), rather than simple features (e.g., specific colors or a prototypical orientation), which may vary too much from one exemplar to the next to guide attention during naturalistic search.

Another dimension of attentional templates, and the topic of this paper, is their spatial specificity. Coming back to our

R. R. Reeder · W. van Zoest · M. V. Peelen (✉)
Center for Mind/Brain Sciences (CIMEC), University of Trento,
Corso Bettini 31, 38068 Rovereto, TN, Italy
e-mail: marius.peelen@unitn.it

street crossing example, we must look out for cars in several possible locations: not just in front of us, but also further down the street or coming from a perpendicular road. Therefore, effective attentional templates would be expected to affect visual processing across a relatively large part of the visual field. Research investigating attention to low-level features (e.g., motion direction or color) has provided ample evidence that feature-based attention effects are spatially global. For example, distractors of a target color can capture attention even if the target always appears in the center of a display and distractors always appear in the periphery, suggesting that preparatory attention for a target color is spread across the visual field (Folk, Leber, & Egeth, 2002). There also is neural evidence that feature-based attention modulates processing across the visual field, as shown by single-cell recordings (Maunsell & Treue, 2006), EEG (Zhang & Luck, 2009), and fMRI (Serences & Boynton, 2007).

We asked whether the attentional template for naturalistic category search similarly influences visual processing in a spatially global way, and whether this effect is involuntary. If this were the case, we would expect that objects matching the template capture attention regardless of where they appear in the display, and even when such capture is detrimental to task performance. In three experiments, subjects were cued to detect people and cars in photographs of real-world scenes presented above and below a central fixation cross. On a subset of trials (dot-probe trials), instead of scenes, silhouettes of people and cars appeared to the left and right of fixation. Subjects were instructed to ignore these stimuli, but to indicate the location of a dot probe that could appear to the left or right of fixation following the silhouettes. This brief (100 ms) dot probe appeared quickly (117 ms) after the onset of the silhouettes. We have previously found that silhouettes of the target category (sharing diagnostic shape features with the target category) capture attention, resulting in faster detection of the dot probe when it was presented in the same location as a template-matching stimulus (consistent trials) than when it was presented in the opposite location (inconsistent trials; Reeder & Peelen, 2013). In that study, silhouettes of the target category reliably captured attention even when scenes and silhouettes appeared in different locations, suggesting that the search template for familiar real-world object categories may affect processing across the visual field.

Using a variant of the attentional blink paradigm, Wyble, Folk, and Potter (2013) similarly provided evidence that distractor images can capture attention when they are of the same conceptual category as the target. Participants in this study were asked to report a target category (e.g., bird, fruit, or vehicle) in a rapid serial visual presentation (RSVP) stream. Just prior to the presentation of the target, distractor images appeared in the periphery above and below the central stream. The results generally showed that the presence of a distractor of the target category hampered performance. More

specifically, in one experiment (Experiment 3), a copy of the central target scene was shown directly after the peripheral same-category distractor scene or at the other possible distractor location. Central task performance was better when the target scene was presented redundantly in the location of the same-category distractor relative to the location of a different-category distractor, indicating spatial orienting to the same-category distractor scene.

It is not clear, however, whether the spatial orienting effects observed in these previous studies (Reeder & Peelen, 2013; Wyble et al., 2013) reflected involuntary attentional capture or instead more strategic orienting mechanisms. For example, subjects could have voluntarily oriented to template-matching stimuli, because it did not necessarily impair performance (targets were presented equally often in the matching and non-matching distractor locations). Furthermore, it is not clear whether these effects reflected capture by attentional templates or instead reflected other processes related to the encoding and maintaining of the category cues, such as semantic priming. In the current study, we sought to explore these possibilities in three experiments.

To investigate the automaticity of the previously observed consistency effect (i.e., faster responses to a dot presented in the location of a silhouette of the cued object compared with a silhouette of the uncued object), in Experiments 1 and 2 we modified the dot-probe task of Reeder & Peelen (2013) to make the location of silhouettes of the uncued category predictive of the location of the subsequent dot-probe. Manipulating the validity of task-irrelevant stimuli in predicting a task-relevant stimulus has been used previously to determine the automaticity of attentional orienting (Jonides, 1981; Downing, Dodds, & Bray, 2004; Spence & Driver, 1994).

In Experiment 3, we investigated whether the observed consistency effects could be explained by non-attentional processes related to encoding and maintaining the cues rather than capture by an attentional template. To rule out such effects, we added a condition in which category cues were used to instruct participants to detect category-related objects (e.g., gear shift for car search) rather than category exemplars themselves. These category-related objects shared no diagnostic visual features with the category silhouettes on dot-probe trials. If the capture effects solely reflected the semantic encoding of the category cues, the silhouettes should capture attention regardless of whether preparatory processes allowed for a diagnostic visual representation of the target category.

Experiment 1

The first experiment was conducted to explore whether stimuli that match the category-level search template capture attention involuntarily. If the search template biases attention in a spatially global way, subjects should be captured by silhouette

stimuli on dot-probe trials even if these appear in search-irrelevant locations and attending to them impairs performance. We tested this prediction by making the target-matching stimuli on dot-probe trials counter-predictive of the location of the dot probe. Dot-probes appeared in the location of the silhouette of the uncued category on 68.75 % of dot-probe trials. Subjects were told before experimentation that paying attention to the cued category on dot-probe trials would harm performance. Furthermore, the search task stimuli and the dot-probe task stimuli were always and consistently presented in different, nonoverlapping locations. If the consistency effect reflects voluntary orienting rather than attentional capture, subjects should be able to ignore the silhouettes altogether. If, however, we observe a consistency effect on dot-probe trials following these instructions, it would provide evidence that the consistency effect reflects involuntary attentional capture by features matching the active search template.

Methods

Subjects

Twelve undergraduate and graduate students from the University of Trento (11 women) participated in the experiment for payment. All subjects had normal or corrected-to-normal visual acuity and were between the ages of 18 and 28 years (mean age = 22.5 years). The research protocol adhered to the ethical principles of the Declaration of Helsinki.

Stimuli

All stimuli were presented on a 19-inch Dell 1905 FP monitor with a screen resolution of 1280 x 1024 pixels and 60-Hz refresh frequency (Dell Inc., Round Rock, TX). Subjects sat 57 cm from the screen. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), a toolbox based on the Psychophysics Toolbox for MATLAB (The MathWorks, Inc., Natick, MA).

A fixation cross and uppercased letter cues appeared centered on the screen in 70-point “strong” Times New Roman font. The fixation cross had dimensions of 31 x 31 pixels

subtending 0.92° in height and width, and letters had dimensions of 70 x 70 pixels subtending 2.1° in height and width.

Probe task stimuli

Stimuli presented in the dot-probe trials (25 % of trials) were composed of 160 upright silhouettes made from photographs of cars (80) and people (80) without scene background (Fig. 1a). Most images were obtained from free-access online image sources and were chosen to encompass a variety of viewpoints and features (e.g., a crouching child, a woman standing, a pickup truck as seen from behind, a sedan as seen from the side). Heads were removed from all images of people to be consistent with the previous behavioral study on which these experiments are based (Reeder & Peelen, 2013).

None of the stimuli presented in the dot-probe trials were shown in the search task. Stimuli could appear in three possible sizes (100 x 100, 180 x 180, or 200 x 200 pixels, or 2.96 x 2.96°, 5.32 x 5.32°, or 5.91 x 5.91° of visual angle, respectively) and in three different screen locations along the X-axis, subtending 6.47°, 7.99°, or 10.76° of visual angle. Size and location values were chosen randomly on each trial and independently for the left and right stimulus. On each dot-probe trial, a single car and person silhouette appeared to the left and right of fixation, and cars and people appeared an equal number of times on either side. Each image was repeated once in an experiment.

Natural scene stimuli

Stimuli presented in the search trials (75 % of trials) were 480 color photographs of real-world scenes obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008; see Fig. 1b for some examples) and were divided into scenes containing cars (n = 240), people (n = 240), both cars and people (n = 240), or neither cars nor people (n = 240). Two scenes appeared on every trial, and no scene was repeated within the experiment.

Scenes were scaled to 548 x 411 pixel resolution, subtending a visual angle of 16.08 x 12.18°. Scenes were

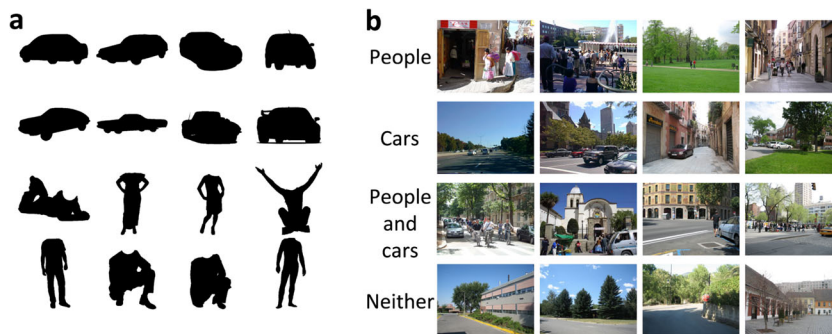


Fig. 1 (a) Examples of attentional capture stimuli (upright silhouettes) and (b) natural scene search stimuli in all experiments

presented 7.41° from the center of the screen to the center of the image, above and below fixation.

Procedure

All subjects ($N = 12$) took part in one practice block followed by 9 blocks of 64 trials each. Each block was made up of two tasks as illustrated in Fig. 2: a search task on 75 % of trials and a dot-probe task on 25 % of trials. Trials were randomized within a block, so subjects could not predict whether the search task or dot-probe task would appear on any given trial.

The search trials and dot-probe trials both started with the presentation of a fixation cross for 500 ms, followed by a single letter for 500 ms: “P” for “persona” or “M” for “macchina” (the Italian words for person and car, respectively). After the letter, another fixation cross appeared for 1 s.

On search trials, subjects would see two photographs of natural scenes for 67 ms, followed by a 350-ms mask.

Subjects were required to respond whether the cued object category appeared above or below fixation using the “up” and “down” arrow keys, respectively. The two scenes that appeared could either be one containing cars and the other containing people, or one containing both cars and people and the other containing no cars and no people. This structure allowed us to present people and cars on every trial without making the presence of one category informative of the absence of the other. Each of the four scene types appeared in each possible location an equal number of times above and below fixation.

On dot-probe trials, following the fixation cross, two images (upright silhouettes of a person and a car) appeared to the left and right of fixation for 67 ms. People and cars were equally likely to appear on the left or right. A 50-ms fixation succeeded the silhouettes, after which a dot probe appeared for 100 ms, 8.52° from the center of the screen on the left or right. Subjects were required to respond whether the dot probe appeared on the left or right of fixation using the “left” and

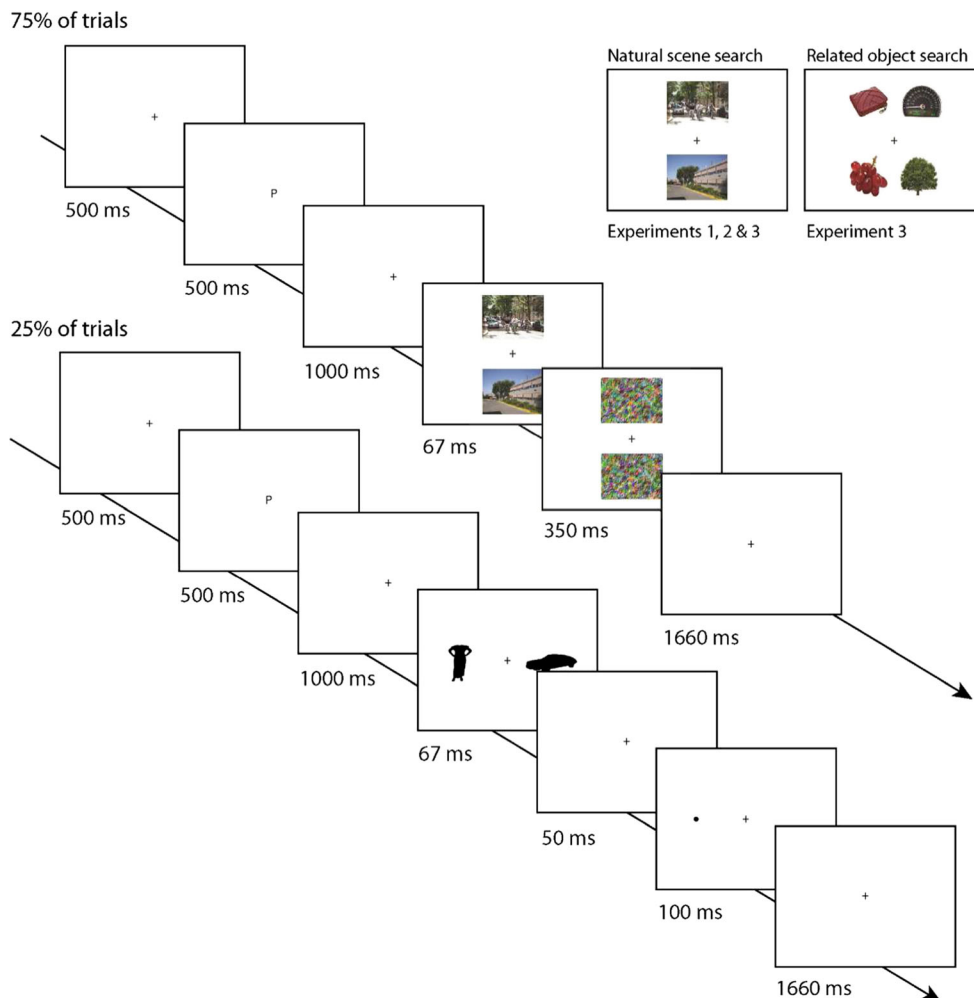


Fig. 2 Schematic of the paradigm used in all experiments. In Experiments 1 and 2 and the natural scene search condition of Experiment 3, natural scenes appeared on search trials (75 % of trials). In the first 5 blocks of Experiment 3, objects related to cars and people

appeared above and below fixation instead of scenes, and subjects were required to detect objects related to the cued category. On dot-probe trials (25 % of trials), silhouettes of cars and people appeared to the left and right of fixation, followed by a dot probe to the left or right of fixation

“right” arrow keys, respectively. They were instructed to ignore the silhouettes that appeared prior to the probe. The proportion of consistent dot-probe trials was 5/16. Consistent trials were those in which the cued category (e.g., the person silhouette following the “P” cue) appeared on the same side of fixation as the dot probe; inconsistent trials were those in which the cued category appeared on the opposite side of fixation. Therefore, the dot probe was inconsistent with the cued category on 68.75 % of trials. Prior to experimentation, subjects were told that paying attention to the cued category on dot-probe trials would be detrimental to performance.

Analysis

We analyzed accuracy and reaction time (RT) for consistent and inconsistent trials in the dot-probe task. Only correct trials were included in the RT analysis. Subjects were excluded from analysis if their mean dot-probe task accuracy fell 2 standard deviations below the group mean for the experiment. One subject was excluded from Experiment 1 based on this criterion. Accuracy results are reported in Table 1 as our main measure of interest was RT, and dot detection accuracy was consistently high with no significant differences between conditions. Cohen’s *d* was used to report the effect sizes for paired samples *t*-tests throughout the paper, and was calculated as the *t* value divided by the square root of the number of subjects.

For the search task, we analyzed accuracy and RT. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 2. The *Results* section reports the RT results of the dot-probe task only.

Results

Figure 3 illustrates mean RT for consistent and inconsistent dot-probe trials of Experiment 1. A paired-samples *t*-test with consistency as the within-subjects factor revealed a strong consistency effect, $t(10) = -3.43, p = 0.006, d = -0.917$, in which subjects responded significantly faster on consistent trials compared with inconsistent trials. Although the RT differences reported here (and indeed, throughout the paper) are on the order of 10-20 ms (similar to previous studies of attentional capture by template-matching probes: e.g., Downing,

Table 1 Accuracy (% correct) for dot-probe trials in Experiments 1, 2, and 3 with standard deviations. Results are reported separately for consistent and inconsistent trials

Experiment	Consistent trials	Inconsistent trials
1	98.3 ± 1.7	97.6 ± 2.7
2	98.6 ± 2.2	98.1 ± 1.8
3 (Natural scenes)	98.4 ± 2.8	98.4 ± 2.5
3 (Related objects)	99.0 ± 1.5	98.2 ± 2.5

Table 2 Reaction time and accuracy results for the search task in Experiments 1, 2, and 3 with standard deviations

Experiment	Reaction time (ms)	Accuracy (% correct)
1	669 ± 172	75.0 ± 9.3
2	699 ± 104	75.6 ± 8.9
3 (Natural scenes)	679 ± 122	83.4 ± 6.7
3 (Related objects)	800 ± 148	79.7 ± 9.5

2000; Reeder & Peelen, 2013), effects were reliable across subjects, as indicated by large effect sizes (partial eta squared and Cohen’s *d* values for ANOVAs and *t*-tests, respectively).

Discussion

Experiment 1 investigated whether subjects were able voluntarily and strategically to withhold orienting to a probe stimulus that was related to an active search template but presented at a search-irrelevant location. In this experiment, the proportion of consistent trials in the dot-probe task was 31.25 % and subjects were told that attending to silhouettes of cued categories would be detrimental to performance. Nevertheless, the results showed that subjects oriented to the silhouettes on dot-probe trials even when it was disadvantageous to performance, suggesting that these consistency effects reflected involuntary attentional capture.

Experiment 2

Experiment 2 was conducted to replicate and extend the findings of Experiment 1. The paradigm was the same as in Experiment 1, with the dot-probe appearing in the location of the silhouette of the uncued stimulus for 11 of 16 (68.75 %) dot-probe trials. The search task continued to be presented on 75 % of trials to ensure the activation of a search template for the target category. In Experiment 2, the task instructions differed from Experiment 1, in that subjects were explicitly

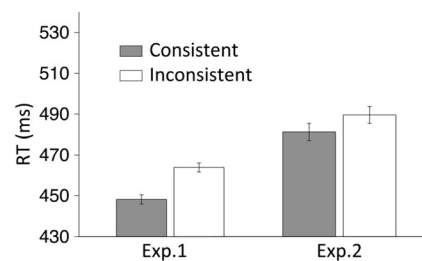


Fig. 3 Reaction time (RT) for consistent and inconsistent trials in the dot-probe task of Experiments 1 and 2. In Experiment 1, subjects received instructions to ignore the silhouettes because attending to them would impair dot-probe RT performance. In Experiment 2, subjects were told to attend to silhouettes of uncued items, because this would improve dot-probe RT performance. Error bars represent the standard error of the mean after adjusting for between-subjects variance (Loftus & Masson, 1994)

informed of the higher probability of a dot probe appearing on the same side of fixation as the silhouette of the uncued category on dot-probe trials. The purpose of these instructions was to induce subjects to switch their orienting strategies between the search task and the dot probe task if at all possible—adopting different templates for different locations. As an additional difference, the category cues in Experiment 2 were changed from letters to arbitrary colors, to remove the intrinsic relationship between the cue and the relevant category for the scene locations. Finding a reliable consistency effect across both Experiments 1 and 2 would increase our confidence that the consistency effects represent an automatic capture effect contingent on a search template spread across the visual field.

Methods

Subjects

There were 14 subjects in Experiment 2 (12 women). All subjects had normal or corrected-to-normal visual acuity and were between the ages of 19 and 33 (mean 24.7) years. The research protocol adhered to the ethical principles of the Declaration of Helsinki.

Stimuli

All stimuli were presented using the same hardware and software as in Experiment 1. Letter cues were replaced with orange and blue circle cues, which appeared centered on the screen, subtending 148 pixels (4.37°) in diameter.

Probe task stimuli

The same probe task stimuli were used in Experiments 1 and 2.

Natural scene stimuli

Natural scene stimuli were taken from the same database and were presented using the same dimensions as in Experiments 1.

Procedure

All subjects (N = 14) except one were different from those who participated in Experiment 1. Experiment 2 was the same as Experiment 1 with two exceptions: first, subjects were cued to search for cars and people with orange and blue colored circles; an orange circle cued cars for half of subjects and cued people for the other half of subjects. Second, prior to experimentation, subjects were explicitly informed that the dot on dot-probe trials would most likely appear on the same side of fixation as the uncued category.

Analysis

We analyzed accuracy and reaction time (RT) for consistent and inconsistent trials in the dot-probe task. Only correct trials were included in the RT analysis. Accuracy results are reported in Table 1, because RT was our measure of interest and there were no significant effects of accuracy, as in Experiment 1. For the search task, we analyzed accuracy and RT. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 2. The *Results* section reports the RT results of the dot-probe task only.

Results

Figure 3 illustrates the mean RT for consistent and inconsistent trials on dot-probe trials of Experiment 2 (N = 14). A paired-samples *t*-test with consistency as the within-subjects factor revealed a significant consistency effect, $t(13) = -3.012$, $p = 0.010$, $d = -0.805$, such that subjects responded faster to dot probes that appeared on the same side of fixation as the cued stimulus.

Discussion

In Experiment 2, we reasoned that if observers are able to modulate voluntarily the use of a search template at different locations, they should be able to prioritize processing of the silhouette of the uncued category on dot-probe trials; nevertheless, the results continued to reveal consistency effects as in Experiment 1, providing further evidence that the category-level search template directs attention automatically to template-matching features presented across the search display, regardless of potentially more beneficial strategies one could use.

Experiment 3

We conducted Experiment 3 to determine whether the observed consistency effects reflect capture by a spatially global search template or, alternatively, non-attentional effects related to encoding and maintaining the preceding category cue (e.g., semantic priming). For example, it cannot be excluded that subjects oriented to the car silhouettes on car search trials because they held the word “car” in mind while preparing for search, without requiring an attentional template of diagnostic car features. Experiment 3 was conducted to investigate whether subjects would orient to silhouettes on dot-probe trials when they were instructed to detect car- or person-related items, which would not require a car- or person-diagnostic template. In five consecutive blocks of trials, instead of searching for cars or people,

subjects were instructed to search for objects semantically related to cars or people. This “related object” search condition required preparation for a variety of objects that are united by their semantic association with the cued category but have virtually no overlapping diagnostic shape features; in other words, a template composed of visual features could not be activated in this condition. Natural scene search (i.e., the search task that was performed in Experiments 1 and 2) occurred in another five consecutive blocks of trials following the related object search condition. Importantly, the cueing procedure and the intermixed dot-probe trials were the same for both search conditions. If the consistency effect on dot-probe trials can be fully explained by general effects of the category cue, we would expect similar consistency effects when subjects searched for category-related objects as when they searched for category exemplars.

Methods

Subjects

Thirteen undergraduate and graduate students from the University of Trento (12 women) participated in the experiments for payment. All subjects had normal or corrected-to-normal visual acuity and were between the ages of 19 and 25 (mean 21.7) years. The research protocol adhered to the ethical principles of the Declaration of Helsinki.

Stimuli

All stimuli were presented using the same hardware and software as in Experiments 1 and 2. The same letter cues were used as in Experiment 1.

Probe task stimuli

The same probe task stimuli were used in Experiments 1-3.

Natural scene stimuli

Natural scene stimuli were taken from the same database and were presented using the same dimensions as in Experiments 1 and 2. Half as many images (240) were presented in the natural scene search condition of Experiment 3 as in the previous experiments and could contain cars ($n = 120$), people ($n = 120$), both cars and people ($n = 120$), and neither cars nor people ($n = 120$).

Car- and person-related object stimuli

Stimuli presented in the search trials (75 % of trials) were color photographs of person-related objects (e.g., wrist-watches, backpacks, sunglasses, hats), car-related objects (e.g., gas cans, GPS, air fresheners, seatbelts), and botanical objects (e.g., fruits, vegetables, trees, flowers; see Fig. 4 for examples and Appendix for full list). Each display was made up of image pairs that consisted of one car-related object and one person-related object ($n = 120$), one car-related object and one botanical object ($n = 120$), one person-related object and one botanical object ($n = 120$), or two botanical objects ($n = 120$). No image pair was repeated within the experiment.

Image pairs were scaled to fit within a boundary that was 548 x 411 pixel resolution subtending a visual angle of 16.08 x 12.18° and were presented 7.41° from the center of the screen to the center of the image pair, above and below fixation.

Procedure

All subjects ($N = 13$) except two were different from those that participated in Experiments 1 and 2. Subjects took part in two



Fig. 4 Examples of the search stimuli in the related object search condition of Experiment 3. Person- and car-related images could either appear together above or below the fixation cross or on opposite sides of fixation from each other. In the former case, two botanical objects would

appear on the opposite side of fixation, and in the latter case, each object would appear next to a botanical object. Botanical objects added clutter to the search arrays to imitate the clutter of natural scenes

separate search conditions: natural scene search and related object search. Whether the natural scene search condition or related object search condition was performed first was counterbalanced across subjects.

Natural scene search condition

All subjects ($N = 13$) took part in one practice block followed by 5 blocks of 64 trials each. Each block was made up of two tasks: a search task on 75 % of trials and a dot-probe task on 25 % of trials, as illustrated in Fig. 2. All search trials in the natural scene search condition corresponded to the search trials presented in Experiments 1 and 2. The dot-probe task followed the same procedure as in the previous experiments with one exception: silhouettes did not predict the dot-probe's location, and the dot appeared on the left and right an equal number of times.

Related object search condition

All subjects ($N = 13$) were first familiarized with the stimuli related to each category and completed one practice block to ensure that they understood the task. All subjects then took part in 5 experimental blocks of 64 trials each. Each block was made up of two tasks: a search task on 75 % of trials and a dot-probe task on 25 % of trials. Trials were randomized within a block so subjects could not predict whether the search task or dot-probe task would appear on any given trial.

The search task and dot-probe task both started with the presentation of a fixation cross for 500 ms, followed by a single letter for 500 ms: "P" for "persona" or "M" for "macchina" (the Italian words for person and car, respectively). After the letter, another fixation cross appeared for 1 s. Dot-probe trials were the same as in the natural-scene search condition.

On search trials, subjects would see four images (2 objects conceptually related to cars and people, and 2 botanical objects) for 67 ms, followed by a 350-ms mask. Two images always appeared above fixation and two appeared below. Subjects were required to respond whether an object conceptually related to the cued category appeared above or below fixation using the "up" and "down" arrow keys, respectively. There were two possibilities for the position of the images: either the car-related object and person-related object could appear on opposite sides of fixation, or they could appear on the same side of fixation. In the first instance, the car-related object would appear side-by-side to a botanical object, and on the opposite side of fixation, the person-related object would appear side-by-side to a botanical object. In the second instance, the car-related object would appear side-by-side to the person-related object, and on the opposite side of fixation, two botanical objects would appear side-by-side. Each of the image types (car-related objects, person-related objects, or

botanical objects) appeared an equal number of times above and below fixation. The car- and person-related objects appeared on the left and right with respect to the other image in the pair an equal number of times.

Analysis

We analyzed accuracy and reaction time (RT) for consistent and inconsistent trials in the dot-probe task of each search condition. Only correct trials were included in the RT analysis. Subjects were excluded from analysis if their mean dot-probe task accuracy fell 2 standard deviations below the group mean for the experiment. One subject was excluded from Experiment 3 based on this criterion. Accuracy results are reported in Table 1; RT was our measure of interest and there were no significant effects of accuracy in this experiment.

We analyzed accuracy and RT for the search task of each search condition. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 2. The Results section reports the RT results of the dot-probe task only.

Results

Figure 5 illustrates mean RT for consistent and inconsistent trials in the dot-probe tasks associated with the natural scene search condition and the related object search condition. A repeated-measures ANOVA with search condition (natural scene, related object) and consistency as factors revealed a significant interaction, $F(1,11) = 8.45$, $p = 0.014$, $\eta_p^2 = 0.434$, indicating that the consistency effect in the natural scene search condition ($t(11) = -3.64$, $p = 0.004$, $d = -1.05$) was significantly greater than the consistency effect in the related object search condition ($t(11) = -0.31$, $p = 0.76$, $d = -0.089$).

Discussion

In Experiment 3, subjects were required to search either for cars and people (natural scene search) or for objects conceptually related to cars and people (related object search). RT analysis revealed a significant consistency effect for dot-probe trials intermixed in the natural scene search condition but not

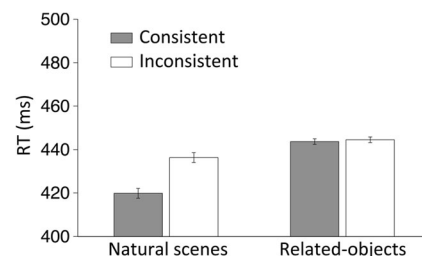


Fig. 5 Reaction time (RT) for consistent and inconsistent trials in the dot-probe tasks of the natural scene and related object search conditions of Experiment 3. Error bars represent the standard error of the mean after adjusting for between-subjects variance (Loftus & Masson, 1994)

for dot-probe trials intermixed in the semantically related object search condition. These results indicate that the consistency effect in the natural scene search condition reflected capture contingent on the attentional template, rather than more general effects of encoding and maintaining the category cues.

General discussion

We tested whether attentional templates activated for detecting real-world object categories automatically and involuntarily affect visual processing across the visual field. The results of Experiments 1 and 2 revealed that attention orienting to search-irrelevant, yet template-matching stimuli is involuntary. Experiment 3 showed that orienting to these stimuli was specific to conditions in which subjects actively maintained a category-diagnostic search template. These findings collectively suggest that orienting to template-matching distractors in this paradigm is due to spatial attentional capture by features matching a spatially unrestricted search template composed of category-diagnostic features.

An important aspect of the current study was that search targets and dot-probe targets were always presented in different locations with no overlap between the scene and probe stimuli. The finding that contingent attentional capture occurs for the dot-probe task stimuli—presented in locations that were always irrelevant for the search task—indicates that the search template affects visual processing across the search display. The results of Experiments 1 and 2 indicate that subjects were unable to inhibit orienting to template-matching stimuli in search-irrelevant locations, suggesting that the spatially global nature of the category-level template is, to some extent, obligatory. Results further showed that subjects were unable to activate different templates for different spatial locations (e.g., attending to cars in the scene task locations and to people in the dot-probe task locations), in line with previous findings that only one template may be active at a time (Houtkamp & Roelfsema, 2009). Together, these results suggest that, at least for category-level search, a single search template affects processing across the visual field.

Experiment 3 revealed attentional capture by silhouettes of cars or people when participants searched for car or person exemplars in scenes, but not when they searched for car- or person-related objects. Thus, searching for objects semantically related to an object category does not necessarily lead to attentional capture by category-diagnostic features. This is not to say that semantic information does not contribute to guiding attention during search: in fact, there is evidence that distractors conceptually related to a target can capture attention when they are presented in search-relevant locations (Moore et al., 2003). However, in that study, attentional

capture by conceptually associated objects was not observed in a dot-probe detection paradigm (Moore et al., 2003). Based on this finding, Moore et al. (2003) suggested that attentional capture by semantically associated objects may initially be spatially unspecific (i.e., not leading to spatial attentional capture), with spatial allocation of attention to these distractors occurring at a later stage that requires more processing time (e.g., through longer presentation durations). Furthermore, it is possible that semantic information may play a larger role in target detection when a visually diagnostic search template would be uninformative for the task, for example when target category exemplars share few visual characteristics (e.g., the category fruit). Importantly, our results show that when targets and semantically related distractors are from separate superordinate categories and are presented in search-irrelevant locations (as in Experiment 3), capture does not occur, suggesting the critical factor in producing attentional capture in our paradigm is the activation of a search template composed of visual features diagnostic of the target category.

Finally, the current results provide a behavioral correlate of a recent fMRI study adopting a similar category-level search task (Peelen et al., 2009). In that study, subjects were cued to search for cars or people in photographs of natural scenes. Following the category cue, four scenes appeared around a central fixation cross (above, below, left, right) and subjects were required to search through either the vertically presented scenes or the horizontally presented scenes. Multivoxel pattern analysis of the fMRI data showed similar category-based attentional modulation for scenes presented at task-relevant and task-irrelevant locations, indicating a spatially global effect of category-based attention, as observed behaviorally here.

The current results provide evidence that category-based attention affects visual processing across the visual field, similar to previously documented effects of feature-based attention. This biasing mechanism may help to support visual search for familiar object categories in naturalistic situations, rapidly orienting spatial attention to locations where categorical targets appear.

Acknowledgments The research was funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012," project "Characterizing and improving brain mechanisms of attention – ATTEND." No conflicts of interest, financial or otherwise, are declared by the authors.

Appendix

Car-related objects:

- Air freshener
- Car battery
- Seatbelt

Jumper cable
 Gas can
 Car engine
 Gas cap
 Gear shifter
 GPS
 Hazard light
 Car jack
 Car key
 Cigarette lighter receptacle
 Car floor mat
 Rearview mirror
 Car door opener
 License plate
 Bicycle rack
 Car radio
 Ice scraper
 Car seat
 Turn signal
 Sunshield
 Tachometer
 Radiator
 Air vent
 Steering wheel
 Windshield wiper

Person-related objects:

Belt
 Bowtie
 Bracelet
 Hairbrush
 Briefcase
 Hair dryer
 Earring
 Eyeglasses
 Hat
 Helmet
 Necklace
 Backpack
 Headphones
 Pocket
 Purse
 Ring
 Scarf
 Shirt buttons
 Cane
 Suspenders
 Sweatband
 Tie
 Umbrella
 Wallet
 Watch
 Zipper

References

- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. doi:10.1146/annurev.ne.18.030195.001205
- Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473.
- Downing, P., Dodds, C., & Bray, D. (2004). Why does the gaze of others direct visual attention? *Visual Cognition*, *11*(1), 71–79.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.
- Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476–1492. doi:10.1037/0096-1523.31.6.1476
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*(5), 741–753. doi:10.3758/BF03194741
- Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research*, *73*(3), 317–326. doi:10.1007/s00426-008-0157-3
- Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi:10.3758/Bf03210951
- Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322.
- Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182–189.
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*, 94–97.
- Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision*, *13*(3), 13. doi:10.1167/13.3.13
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1-3), 157–173. doi:10.1007/S11263-007-0090-8
- Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*, *43*(4), 1194–1201. doi:10.3758/s13428-011-0106-8
- Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312. doi:10.1016/j.neuron.2007.06.015
- Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 555–574.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. doi:10.1080/13506280500195250
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687. doi:10.1038/nm870
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 419–433.

Wyble, B., Folk, C., & Potter, M. C. (2013). Contingent attentional capture by conceptually relevant images. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 861.

Zhang, W., & Luck, S.J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, 12(1), 24–25. doi:[10.1038/nn.2223](https://doi.org/10.1038/nn.2223)